

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:54 ; Search time 91.75 Seconds

(without alignments)  
16.543 Million cell updates/sec

Title: US-09-331-631a-8\_COPY\_33\_79

Perfect score: 275  
Sequence: 1 GDDPPKRYEDCRRRCCEMDT.....QCESCKSQYGEKDDQQRHR 47

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	588	1 VCLB_GOSHI	P09801 gossypium h
2	140	50.9	605	1 VCLB_GOSHI	P09799 gossypium h
3	75	27.3	47	1 AGRP_LOFCY	P56768 lufia cyllin
4	67.5	24.5	471	1 VSWA_TFRYBR	P20946 trypanosoma
5	62	22.5	360	1 MJDL_HUMAN	P54252 homo sapien
6	61	22.2	189	1 GRPE_HELPY	P55970 heliobacte
7	61	22.2	1905	1 TAGB_DICDI	P54683 dictyostei
8	59.5	21.6	573	1 GBLI_MAIZE	P15590 zea mays (m
9	59	21.5	1339	1 DPOA_TFRYBR	P27727 trypanosoma
10	58.5	21.3	443	1 FBLA_CRIGR	O55058 cricetus
11	58.5	21.3	543	1 MIG1_KLUMA	P52288 kluyveromyc
12	58.5	21.3	1085	1 YAPA_SCHPO	O09863 schizosacch
13	58	21.1	338	1 LEBB_PEA	P14594 pisum sativ
14	57.5	20.9	443	1 FBL4_HUMAN	O95967 homo sapien
15	57.5	20.9	443	1 FBL4_HUMAN	O94719 mus musculu
16	57.5	20.9	461	1 KLB2_ECOLI	P52605 escherichia
17	57.5	20.9	650	1 Y096_MYCGE	P47342 mycoplasma
18	57	20.7	1898	1 TRHY_HUMAN	O07283 homo sapien
19	56	20.4	301	1 PHLP_RAT	O63737 rattus norv
20	56	20.4	481	1 KNR1_DROVI	O24753 drosophila
21	56	20.4	1816	1 LMA4_HUMAN	O16363 homo sapien
22	55.5	20.2	2327	1 CCAB_MOUSE	O05017 mus musculu
23	55.5	20.2	2336	1 CCAB_MOUSE	O02294 rattus norv
24	55.5	20.2	2339	1 CCAB_MOUSE	O00975 homo sapien
25	55.5	20.2	2339	1 CCAB_MOUSE	O05152 mus musculu
26	55	20.0	462	1 U2R2_MOUSE	O63377 mus musculu
27	55	20.0	540	1 PNR_DROME	P52168 drosophila
28	55	20.0	812	1 PLMN_MOUSE	P20918 mus musculu
29	55	20.0	3344	1 POLG_PRSVH	O01901 p genome po
30	54.5	19.8	707	1 DREB_RAT	O07266 rattus norv
31	54.5	19.8	841	1 IEB3_MCMVS	O69154 murine cyto
32	54	19.6	1744	1 TANA_XENLA	O01550 xenopus lae
33	54	19.6	2254	1 CCAG_RAT	O54898 rattus norv

34	53.5	19.5	405	1 IF4A_CRYPV	O02494 cryptospori
35	53.5	19.5	846	1 ITBX_DROME	P11584 drosophila
36	53.5	19.5	1391	1 MSF2_DROXY	O08696 drosophila
37	53	19.3	60	1 PHLP_HUMAN	O13371 homo sapien
38	53	19.3	342	1 CISP_SCHJA	P43157 schistosoma
39	53	19.3	400	1 T2PB_YEAST	P41896 saccharomyc
40	53	19.3	466	1 CYP8_CAEEL	P52016 caenorhabdi
41	53	19.3	691	1 CNG1_CANFA	O28279 canis famli
42	53	19.3	892	1 BN14_YEAST	P53858 saccharomyc
43	53	19.3	1549	1 TRHY_SHEEP	P22793 ovis aries
44	52.5	19.1	394	1 YEL8_YEAST	P39943 saccharomyc
45	52.5	19.1	538	1 GEL1_YEAST	O12315 saccharomyc

## ALIGNMENTS

RESULT 1					
ID	VCLB_GOSHI	STANDARD:	PRT:	588 AA.	
AC	P09801:				
DT	01-MAR-1989 (Rel. 10, Created)				
DT	01-MAR-1989 (Rel. 10, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
OC	[1]				
RA	SEQUENCE FROM N.A.				
RP	'Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and				
RT	germination. XVIII. CDNA and amino acid sequences of the members of				
RT	the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN				
CC	BODIES.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,				
CC	CONVICILIN, CONGLICININ, ETC.).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb.ch">license@isb.ch</a> ).				
CC	-----				
DR	EMBL: M16891; AAA33071.1; -.				
DR	PIR: A30838; FWCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; -.				
DR	PFAM: PF00546; Seedstore_7s; 1.				
KW	Seed storage protein; signal.				
FT	SIGNAL				
FT	CHAIN				
FT	SEQUENCE				
FT	588 AA; 69729 MW; 636699B29AB8ADB CRC64;				
SO	SEQUENCE				
Query Match		100.0%;	Score 275;	DB 1;	Length 588;
Best Local Similarity		100.0%;	Pred. No. 1,4e-22;		
Matches	47;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
OY	1 GDDPPKRYEDCRRRCCEMDTGGQKEDQOQCESCKSQYGEKDDQQRHR 47				
DB	33 GDDPPKRYEDCRRRCCEMDTGGQKEDQOQCESCKSQYGEKDDQQRHR 79				
RESULT 2					
ID	VCLB_GOSHI	STANDARD:	PRT:	605 AA.	
ID	VCLB_GOSHI				

```

AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chian C.A., Borrito K., Kamalay J.A., Dure L. III:
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001133.
DR PFM: PF00546; Seedstore.75; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;
SQ
Query Match 50.9%; Score 140; DB 1; Length 605;
Best Local Similarity 54.5%; Pred. No. 4.1e-08;
Matches 24; Conservative 11; Mismatches 7; Indels 2; Gaps 1;
QY 2 DDDPKRYEDCRRRCMDTRGQKEOQCEESCKSQYGEKDOQR 45
DB 34 EDDPQRYEDCKRKRCOLRTROTEDQKCDRSETQL--KEEQQR 75
RESULT 3
AGRP_LUFCY STANDARD: PRT: 47 AA.
ID AGRP_LUFCY
AC P56568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 6.5 KDa ARGININE/GLUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).
OS Lufia cylindrica (Smooth loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Lufia.
RN [1]
RP SEQUENCE.
RP TISSUE=SEED;
RX MEDLINE=97357433; PubMed=9214759;
RA Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.:
RT "Primary structure of 6.5K-arginine/glutamate-rich polypeptide from
RT the seeds of sponge gourd (Lufia cylindrica).";
RL Biosci. Biotechnol. Biochem. 61:984-988(1997).
CC -1- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CC -1- MASS SPECTROMETRY: MW=5693.39; METHOD=MALDI.

```

```

CC -1- SIMILARITY: SOME: NO 7S SEED STORAGE PROTEINS.
KW Seed storage protein.
FT DISULFID 12 33
FT DISULFID 16 29
SQ SEQUENCE 47 AA; 5698 MW; 588B0EC82273AC05 CRC64;
Query Match 27.3%; Score 75; DB 1; Length 47;
Best Local Similarity 32.4%; Pred. No. 0.035;
Matches 12; Conservative 10; Mismatches 15; Indels 0; Gaps 0;
QY 5 PPKRYEDCRRRCMDTRGQKEOQCEESCKSQYGEKD 41
DB 5 PRTYEAQCRVQVAHGVGRQRCQVCEKRLRRE 41
RESULT 4
VSWA_TRYBR STANDARD: PRT: 471 AA.
ID VSWA_TRYBR
AC P20946;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE VARIANT SURFACE GLYCOPROTEIN WRATAT A PRECURSOR (VSG).
OS Trypanosoma brucei rhodesiense.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=WRATAT 1;
RX MEDLINE=90290520; PubMed=2357229;
RA Reddy L.V., Hall T., Donelson J.E.:
RT "Sequences of three VSG mRNAs expressed in a mixed population of
RT Trypanosoma brucei rhodesiense.";
RL Biochem. Biophys. Res. Commun. 169:730-736(1990).
CC -1- FUNCTION: VSG FORMS A COAT ON THE SURFACE OF THE PARASITE. THE
CC A SERIES OF ANTIGENICALLY DISTINCT VSGS FROM AN ESTIMATED 1000
CC VSG GENES.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC A SOLUBLE FORM IS RELEASED FROM RUPTURED CELLS BY THE ACTION OF A
CC PI-PLC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M33823; AAA30316.1; -
DR PIR: A35480; A35480.
KW Glycoprotein; Antigen; Trypanosomiasis; GPI-anchor; Membrane; Signal.
FT SIGNAL 1 18
FT CHAIN 19 454
FT PROPEP 455 471
FT LIPID 454 454
FT CARBOHYD 61 61
FT CARBOHYD 133 133
FT CARBOHYD 449 449
SQ SEQUENCE 471 AA; 50537 MW; 57ECC0821042376D CRC64;
Query Match 24.5%; Score 67.5; DB 1; Length 471;
Best Local Similarity 31.0%; Pred. No. 1.8e-16;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;
QY 3 DDDPKRYEDCRRRCMDTRGQKEOQCEESCKSQYGEKDOQ 44
DB 402 NDKAKETE-CNSPKWMDKEKDEKRRKLSBEGKAKEKNE 442
RESULT 5

```

MJD1	HUMAN			PRT:	360 AA.
ID	_MJD1_HUMAN	STANDARD;			
AC	p5A252:				
DT	01-OCT-1996 (Rel. 34,				
DT	01-OCT-1996 (Rel. 34,	Last sequence update)			
DT	01-OCT-1996 (Rel. 34,	Last annotation update)			
DE	MACHADO-JOSEPH DISEASE PROTEIN 1.				
CN	MJD1 OR MJD.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
NC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RP	TISSUE=BRAIN.				
RX	MEDLINE=95179166; PubMed=7874163;				
RA	Kawaguchi Y., Okamoto T., Tanitaki M., Aizawa M., Inoue M.,				
RA	Katayama S., Kawakami H., Nakamura S., Nishimura M., Akiguchi I.,				
RA	Kimura J., Narumiya S., Kakizuka A.;				
RT	"CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1."				
RL	Nat. Genet. 8:221-228(1994).				
CC	-I- POLYMERISM: THE POLY-GLN REGION OF THE MACHADO-JOSEPH PROTEIN IS HIGHLY POLYMORPHIC (14 TO 40 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO ABOUT 68-82 REPEATS IN MJD PATIENTS. LONGER EXPANSIONS RESULT IN EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE DISEASE.				
CC	-I- DISEASE: DEFECTS IN MJD ARE THE CAUSE OF MACHADO-JOSEPH DISEASE (AMD), A NEURODEGENERATIVE DISORDER CHARACTERIZED BY CEREBELLAR ATAXIA, PYRAMIDAL AND EXTRAPYRAMIDAL SIGNS, PERIPHERAL NERVE PALSY, EXTERNAL OPHTHALMOPLEGIA, FACIAL AND LINGUAL FASCICULTATION AND BULGING. THIS DISEASE IS AUTOSOMAL AND DOMINANT, WITH A LATE ONSET OF SYMPTOMS, GENERALLY AFTER THE FOURTH DECADE.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; S75313; AAB33571.1; -.				
DR	WM; 109150; -.				
DR	INTERPRO; IPRO02950; -.				
DR	PFAM; PF02099; Josephin; 1.				
DR	PRINTS; PR01233; JOSEPHIN.				
DR	POLYPHORMISm: Triplet repeat expansion.				
FT	DOMAIN 292 .. 317 POLI-GLN.				
SQ	SEQUENCE 360 AA; 41531 MW; AD6774F29A6CD95D CRC64;				
<hr/>					
	Query Match	22.5%;	Score 62;	DB 1;	Length 360;
	Best Local Similarity	35.0%;	Pred. No. 5.5;		
	Matches 14;	Conservative 12;	Mismatches 12;	Indels 2;	Gaps 1;
QY	10 EDCCRRCR--WDTRGRQRQQQCCESSCKSQDYEEKDDQQRNR 47  ::  :::  ::  ::  ::  ::  ::  ::  ::  ::  ::  :  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  :  Db 279 ELKRKREAYEFKKQDKQQQQQQQQQQQQQQQQQQQQQQRR 318				
<hr/>					
RESULT	6				
ID	GRPPE_HELPHY	STANDARD;	PRT;	189 AA.	
AC	p55970:				
DT	01-NOV-1997 (Rel. 35,	Created)			
DT	01-NOV-1997 (Rel. 35,	Last sequence update)			
DT	30-MAY-2000 (Rel. 39,	Last annotation update)			
DE	GREP PROTEIN (HSP-70 COFACTOR).				
CN	GREP OR HP0110.				
CS	Helicobacter pylori (Campylobacter pylorici).				
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;				
CC	Helicobacter				
NN	[1]				

```

PC SEQUENCE FROM N.A.
RP STRAIN=26695 / ATCC 700392:
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weldman J.M., Fujii C., Bowman C., Watney L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "the complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -I- FUNCTION: STIMULATES, JOINTLY WITH DNAJ, THE ATPASE ACTIVITY OF
CC DNAJ. HELPS TO RELEASE ADP FROM DNAJ THUS ALLOWING DNAJ TO RECYCLE
CC MORE EFFICIENTLY (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE GRPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000532; AAD07179.1; -.
CC TIGR: HP0110; -.
DR INTERPRO: IPR000740; -.
DR PfAM: PF01025; GRPE; 1.
DR PRINTS: PRO0773; GREPEPROTEIN.
DR PROSITE: PS01071; GRPE; 1.
KW Chaperone; Heat shock.
SO SEQUENCE 189 AA; 22040 MW; 62C0991B684F0C2D CRC64;

Query Match 22.2%; Score 61; DB 1; Length 189;
Best Local Similarity 44.4%; Pred. No. 3.9;
Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 17 EWDTRGKREQQQCEESCKSYGKDDQ 43
   1 11 11:11 1 1:
Db 8 EHDHLSKREPECEKACEKQGYEKKOE 34

RESULT 7
TAGB_DICDI STANDARD; PRT; 1905 AA.
ID TAGB_DICDI
AC P54683;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PRSTALM-SPECIFIC PROTEIN TAGB PRECURSOR (BC 3.4.21.-).
GN TAGB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostelida; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AX4;
RX MEDLINE=95262903; PubMed=7744252;
RA Shaulsky G., Kuspa A., Loomis W.F.;
RT "A multidrug resistance transporter/serine protease gene is required
RT for prestalk specialization in Dictyostelium."
RL Genes Dev. 9:1111-1122(1995).
CC -I- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
CC INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC -I- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC -I- SIMILARITY: STRONG, TO TAGC.

```

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U20432; AAA62212.1; -  
DR HSSP: P13569; INBD.  
DR DICTYDB: DD02059; TAGB.  
DR INTERPRO: IPR000209; -  
DR INTERPRO: IPR001140; -  
DR INTERPRO: IPR001617; -  
DR PFAM: PF00664; ABC\_membrane; 1.  
DR PFAM: PF00005; ABC\_tran; 1.  
DR PRINTS: PR00723; SUBTILISIN.  
DR PROSITE: PS00136; SUBTILASE\_ASP; FALSE\_NEG.  
DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
KW Hydroxylase; Serine protease; ATP-binding; Transport; Transmembrane;  
Signal.  
FT SIGNAL 1 ? POTENTIAL.  
FT CHAIN ? PRESTALK-SPECIFIC PROTEIN TAGB.  
FT DOMAIN 378 700 PROTEASE.  
FT TRANSMEM 1011 1031 ABC\_TRANSPORTER.  
FT TRANSMEM 1076 1096 POTENTIAL.  
FT TRANSMEM 1121 1141 POTENTIAL.  
FT TRANSMEM 1210 1230 POTENTIAL.  
FT TRANSMEM 1309 1329 POTENTIAL.  
FT TRANSMEM 1332 1352 POTENTIAL.  
FT ACT\_SITE 387 387 CHARGE\_RELAY\_SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 432 432 CHARGE\_RELAY\_SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 695 695 CHARGE\_RELAY\_SYSTEM (BY SIMILARITY).  
FT NP\_BIND 1553 1560 ATP (POTENTIAL).  
FT DOMAIN 63 67 POLY-GLN.  
FT DOMAIN 95 104 POLY-ASN.  
FT DOMAIN 107 134 POLY-ASN.  
FT DOMAIN 311 321 POLY-SER.  
FT DOMAIN 833 839 POLY-SER.  
FT DOMAIN 838 844 POLY-GLY.  
FT DOMAIN 871 876 POLY-LEU.  
FT DOMAIN 1012 1015 POLY-ILE.  
FT DOMAIN 1386 1389 POLY-GLY.  
FT DOMAIN 1398 1404 POLY-GLY.  
FT DOMAIN 1445 1450 POLY-ASN.  
FT DOMAIN 1765 1779 POLY-ASN.  
FT DOMAIN 1782 1785 POLY-SER.  
FT DOMAIN 1807 1812 POLY-PRO.  
FT DOMAIN 1815 1860 POLY-GLN.  
FT DOMAIN 1872 1878 POLY-PRO.  
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1172 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1522 1522 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1558 1558 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

Query Match 22.2%; Score 61; DB 1; Length 1905;  
Best Local Similarity 28.3%; Pred. No. 33;  
Matches 13; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

OY 2 DDDPKRYEDCRRRCMDTRGKQEQCEESCKSOYGEKDQOQRH 47  
DB 1805 DDPKRRPPQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 1850

RESULT 8  
G1B1 MAIZE  
ID G1B1 MAIZE STANDARD; PRT; 573 AA.  
AC P15590;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLOBULIN-1 S ALLELE PRECURSOR (G1B1-S) (7S-LIKE).  
GN G1B1.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. INbred LINE VA26;  
RA Belanger F.C., Kriz A.L.;  
RT "Molecular characterization of the major maize embryo globulin encoded  
RT by the G1B1 gene."  
RL Plant Physiol. 91:636-643(1989).  
[2]  
RP SEQUENCE OF 87-100.  
RX MEDLINE=89374022; PubMed=2775172;  
RA Kriz A.L.;  
RT "Characterization of embryo globulins encoded by the maize G1B  
RT genes."  
RL Biochem. Genet. 27:239-251(1989).  
CC -I- PM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE  
CC MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.  
CC -I- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING G1B1 ALLELES HAVE  
CC THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL  
CC PROTEINS, RESPECTIVELY.  
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
CC CONVICILIN, CONGLICANIN, ETC.).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M24845; AAA33467.1; -  
DR HSSP: P50477; ICAM.  
DR MAIZEDB: P0181; -  
DR INTERPRO: IPR001113; -  
DR PFAM: PF00546; Seedstore\_7s; 1.  
KW Seed storage protein; Signal.  
FT SIGNAL 1 18 OR 21 (POTENTIAL).  
FT PROPEP 19 86  
FT CHAIN 87 573 GLOBULIN-1 S ALLELE.  
FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 573 AA; 65029 MW; 525ED1D00A062976 CRC64;

Query Match 21.6%; Score 59.5; DB 1; Length 573;  
Best Local Similarity 38.9%; Pred. No. 16;  
Matches 14; Conservative 7; Mismatches 12; Indels 3; Gaps 2;

OY 12 CRRCEMDTRGKQEQCEESCKSOYGEKDQOQRH 46  
DB 40 CVRCE--DRPHQRPCLCEQCEERERKROERSRH 73

RESULT 9  
DPOA-TRYBB  
ID DPOA-TRYBB STANDARD; PRT; 1339 AA.  
AC P27727;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA POLYMERASE ALPHA CATALYTIC SUBUNIT (EC 2.7.7.7) (DNA POLYMERASE







